A simple conceptual framework and nomenclature for studying repeated, parallel and convergent evolution

José Cerca[1,*]

1 - Department of Natural History, NTNU University Museum, Norwegian University of Science and Technology, Trondheim, Norway

* correspondence at jose.cerca@gmail.com

A simple conceptual framework and nomenclature for studying repeated, parallel and convergent evolution

Abstract

Parallel and convergent evolution are textbook examples of the role of natural selection in evolution. However, these terms are used interchangeably, and sometimes with conflicting meanings. This has resulted in confusion, which hampers the understanding of the processes underlying these important forms of evolution. In this synthesis, I discuss the issues with current definitions of parallel, repeated and convergent evolution, and provide a framework aimed at solving these issues. This framework makes an important distinction between environmental properties and organismal properties, with the first involving the role of similar and non-similar environmental and selective pressures. The organismal properties include the genomic-basis where the process (mutation, standing genetic variation

and gene flow) and the location (homologous nucleotide, homologous gene region or non-homologous gene region) are emphasised, and the phenotype (convergent or parallel evolved). I restrict the use of the terms parallel (evolution of similar and derived phenotypes, from similar ancestral phenotypes) and convergent (evolution of similar and derived phenotypes, from dissimilar ancestral phenotypes) evolution to the phenotypic level, thereby restoring its original meaning before the genomic revolution. I argue that this framework and nomenclature provide a clear resolution to study parallel and convergent phenotypic evolution across fields while maintaining the interest in its genomic and ecological grounds. Crucially, this framework stresses the importance of a multidisciplinary focus, integrating ecology, genomics, and phenotypes to determine whether parallel or convergent phenotypic evolution has taken place.

Introduction

Convergent, parallel, and repeated phenotypic evolution (BOX 1) illustrate the seminal role of natural selection in evolution, and have a central role in understanding deterministic outcomes in evolution (Wood *et al.*, 2005; Losos, 2011). The independent and repeated evolution of phenotypes is ubiquitous, occurring between closely related species, such as Galápagos' finches, cichlid fishes and *Anolis* lizards (Losos, 2010; Elmer & Meyer, 2011), and among distantly related lineages such as the evolution of wings in bats, insects, pterodactyls and birds (Stern, 2013). A central challenge of biology then becomes determining the drivers

(why), the underlying mechanisms (how), the rates (how much), and the taxa (who), underlying the repeated evolution of phenotypes (Rosenblum *et al.*, 2014).

Considering the essential role of convergent and parallel phenotypic evolution in shaping our knowledge of natural selection and evolution in general, one would expect that these terms were clearly defined (Stayton, 2015). Surprisingly though, biologists have employed these terms with a variety of meanings (Rosenblum *et al.*, 2014), and it has been argued that the lack of resolution likely stems from the ever-growing contribution of genomic sequencing, which has transformed biology (Wood *et al.*, 2005). As an example, scoring the top 100 cited publications since 2010 and the latest 100 publications after the terms 'convergent evolution' and 'parallel evolution' retrieved multiple definitions, which can be broadly organised into ***(i)*** phenotype-based (parallel or convergent evolution of phenotypic traits), ***(ii)*** genomic-based (parallel or convergent evolution of nucleotide or amino acid sequences), ***(iii)*** evolutionary distance-based (where parallel evolution occurs between closely related lineages and convergent evolution between distantly related lineages), and ***(iv)*** convergence in community composition (community ecology; Supplementary Table 1). Some of the analysed papers used multiple definitions under this classification (Supplementary Table 1).

The lack of consistent definitions in science leads to confusion and unnecessary disagreement. First, patterns and processes are distinct at the genotypic and phenotypic levels, and casually applying terms at different levels muddies the waters (Arendt & Reznick, 2008).

Second, vague or imprecise terms are likely to impact early career researchers and newcomers to the field, who are sedimenting their knowledge, creating a superfluous barrier to their understanding. For instance, in biodiversity genomics parallel evolution has been used to refer to 'parallel changes at the molecular sequence level' (Natarajan *et al.*, 2015), 'parallel replacements/substitutions' (Natarajan *et al.*, 2015; Mendes *et al.*, 2016; Lee *et al.*, 2018), 'parallel evolution and fixation of mutations' (Stern, 2013), 'parallel selection on standing genetic variation' (Pease *et al.*, 2016), 'parallel adaptation' (Stoltzfus & McCandlish, 2017; Bohutínská *et al.*, 2021; Konečná *et al.*, 2021; Szukala *et al.*, 2022), 'parallel evolution of phenotypes' (Colosimo *et al.*, 2005; Szukala *et al.*, 2022), among other terms. Third, because nuances in the terms are likely field-specific (when considering developmental biology, physiology, ecology, genomics as different fields within biology), they may obstruct multidisciplinary efforts. Fourth, discordant definitions are not merely semantic disagreements (Harrison, 2012), as they impact on how we reflect concepts, models, frameworks, patterns, and processes (Harrison, 2012; Stayton, 2015). The language we use, intentionally or inadvertently, guides hypothesis testing, how we think, and the training of future generations (Harrison, 2012). These nuances may also lead to disagreements during evaluation processes such as peer review or grant assessments, as different scientists will have different interpretations for a given term, and thereby result in unnecessary barriers and arguments. In sum, nuances in our definitions, no matter how subtle, can lead to considerably different interpretations and ultimately conclusions.

In one of the most influential attempts to deal with these issues (>500 citations), Arendt and Reznick (2008) suggested dropping the term parallel evolution and only adopting convergent evolution. While this is tempting, the terms parallel and convergent make an important distinction: that a given derived phenotype can evolve from similar (parallel) or dissimilar (convergent) starting points (Leander, 2008). This distinction is fundamental to understanding phenotypic evolution, and begs the integration of the genetic and ecological bases underlying the repeated evolution of phenotypes. This distinction is crucial, and a quick search on ISI web of knowledge shows that even after Arendt and Reznick (2008), parallel evolution has remained a considerably more popular term, being ~4-5x more used, than convergent evolution .

With this in mind, I propose a simple conceptual framework (Figure 1) aimed at solving the inconsistencies in the definitions of convergent, parallel, and repeated evolution. This framework is split into two major components (Wellborn & Langerhans, 2015): the environmental and the organismal properties, with the latter being the genomic basis and the phenotypic basis (Figure 1). The focus on environmental and organismal components highlights the multidisciplinary nature of working with parallel and convergent, which requires the integration of ecological, phenotypic and genomic data.

BOX 1 - definitions

Parallel evolution – evolution of similar and derived phenotypes, from similar ancestral phenotypes.

Convergent evolution – evolution of similar and derived phenotypes, from dissimilar ancestral phenotypes.

Repeated evolution - broad term determining any evolutionary pattern or process that has occurred multiple times (repeatedly). This term should be used in ambiguous cases and it can thus broadly encompass: (i) parallel evolution at the phenotypic level; (ii) convergent evolution at the phenotypic level; (iii) repeated mutations at the same nucleotide or gene; (iv) repeated introgression of alleles; (v) repeated recruitment of standing genetic variation; (vi) similar selective pressures at the environmental domain.

Mutation - change in a DNA sequence (not necessarily restricted to a point-mutation).

Gene flow - Genetic exchange between different lineages (species or populations).

Shared ancestral polymorphism (SAP) or Standing Genetic Variation (SGV) - Genetic variation that has been previously tested by selection and which is common to a group of populations or species.

Mutation at a homologous nucleotide - Mutation in the same nucleotide.

Mutation at homologous loci - Mutation occurring in different nucleotides, in the same gene.

Mutation at non-homologous loci - Mutation occurring in different nucleotides in different genes.

/end **BOX 1**

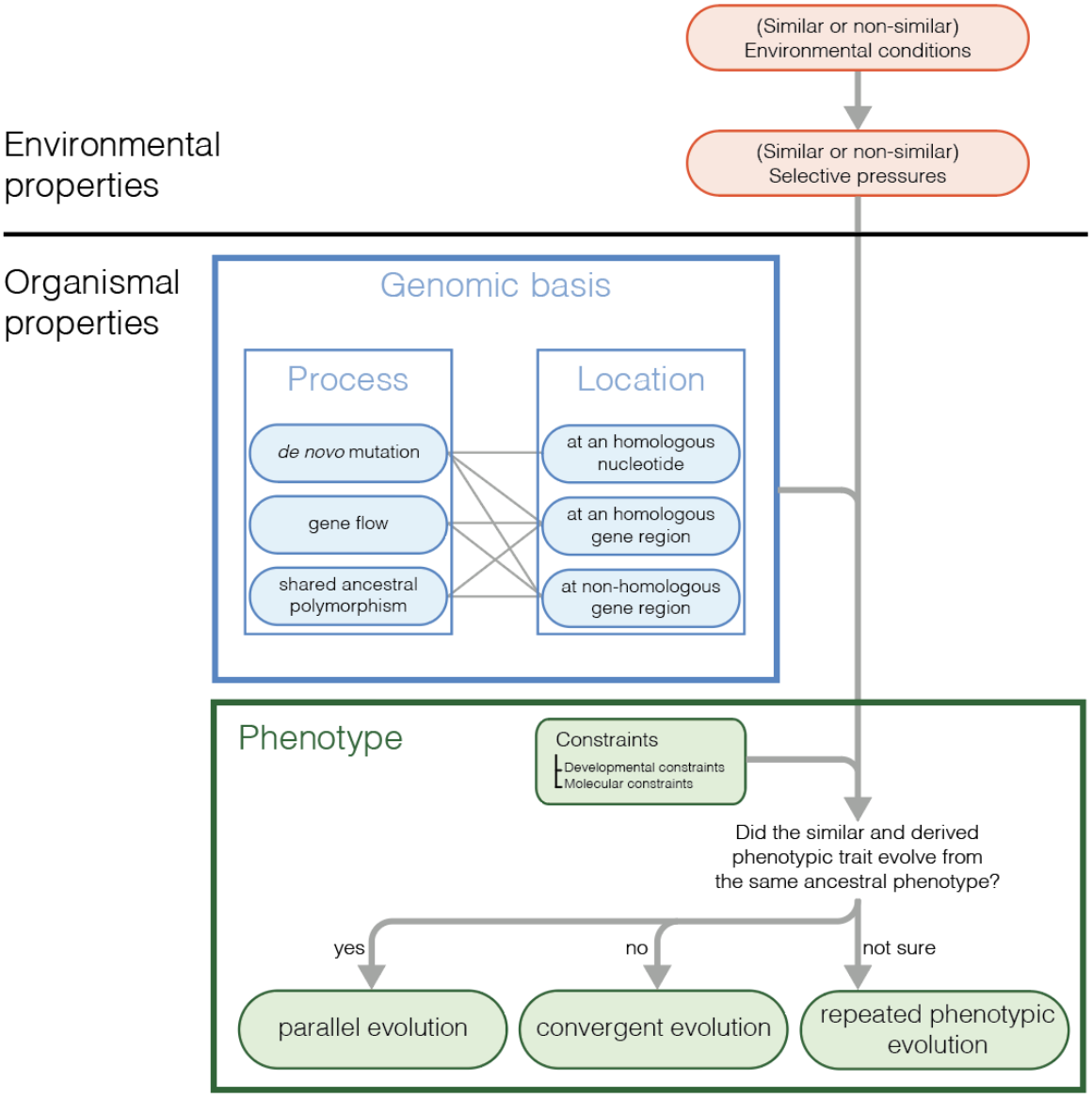

Environmental properties
(Similar or non-similar) Environmental conditions
(Similar or non-similar) Selective pressures
Organismal properties
Genomic basis
Process
Location
de novo mutation
gene flow
shared ancestral polymorphism
at an homologous nucleotide
at an homologous gene region
at non-homologous gene region
Phenotype
Constraints
Developmental constraints
Molecular constraints
Did the similar and derived phenotypic trait evolve from the same ancestral phenotype?
yes
no
not sure
parallel evolution
convergent evolution
repeated phenotypic evolution

Figure 1 Framework for studying parallel and convergent phenotypic evolution. This framework is divided into two major components: the environmental and organismal properties. The first includes environmental conditions, which can be similar or not, and selective pressures. The organismal properties are divided into two elements: the genomic basis and the phenotypic basis. The genomic basis is subdivided into process, including three non-exclusive (see Table 1) and potentially repeatable processes (shared ancestral polymorphism or standing genetic variation, gene flow, and *de novo* mutation), and the location (at the same nucleotide, gene-locus or different genes). *De novo* mutation is the only process that should be considered at all the locations, while standing genetic variation and gene flow should be considered at homologous gene regions and non-homologous gene regions. The terms parallel and convergent evolution strictly apply to the phenotype, which can be influenced by constraints.

Issues of current definitions and justification for the framework

The framework (Figure 1) limits the terms parallel and convergent evolution to the phenotype. The choice of the phenotypic focus is grounded on various reasons.

(i) Definitions based on evolutionary distance lack clarity as they are arbitrary (Arendt & Reznick, 2008). Under this set of definitions, parallel evolution occurs between closely related species, while

convergent evolution occurs between distantly related species. The issue with this definition is that the limits where parallel and convergent evolution occur is relative and thus arbitrary. Consider, for instance, the three-spined stickleback (*Gasterosteus aculeatus*) and the manatee (*Trichechus* spp.), and their common genetic basis for a pelvic-reduced phenotype (Leander 2008). While one may consider that convergence applies between a fish and a mammal, one may also consider this as parallel evolution relative to the phenotypic convergence between eukaryotic microbial lineages with > 950 million years of divergence (Leander 2008).

(ii) Genomic-based definitions (parallel or convergent evolution of genetic sequences or pathways) fail to capture the complexity of genomic patterns and processes because they do not explicitly distinguish between the genomic location nor the processes involved (see discussion below on the genotype). This has led to the 'reinvention of the wheel', where operational concepts with slight nuances are used by different researchers (Supplementary Table 1). For example, some consider parallel evolution to occur when derived phenotypes have a share genetic-origin (e.g. the phenotypes are conferred by mutations in the same gene or genetic pathway), while convergent evolution involves a non-shared genetic-origin (e.g. mutations on distinct genes or pathways). This definition is problematic because, first, delineating what constitutes a pathway remains extremely difficult for non-model organisms (Losos, 2011). Second, because it leads to redundancy. For instance, if we consider a case where three lineages (A, B, and C) repeatedly evolved a derived phenotype. On lineage A the derived

phenotype is conferred by genes 1 and 2, while on lineage B the derived phenotype is conferred by genes 1 and 3, and on lineage C the phenotype is conferred by genes 3 and 4. This would mean that A vs B and B vs C evolved in parallel. However, A vs C would be a case of convergent evolution. It is often the case that the genomic origin involves a complex genetic basis, where a mix of shared and non-shared genetic mechanisms may underpin a phenotype in different lineages (Table 1; see below on stickleback). Others consider that parallel evolution occurs when different mutations occur in different lineages, and these confer the same derived phenotype (Konečná *et al.*, 2021). Strictly speaking, one may argue that two similar and derived phenotypes may not be considered as parallel nor convergent evolution if gene flow or standing genetic variation occur. In sum, current genomic-based definitions may be problematic, and cause unnecessary confusion.

To solve these issues I propose restricting the use of parallel and convergent evolution to the phenotype, and encourage the use of different nomenclature at the environmental-level, and at the genotype-level.

Environmental conditions and selective pressures

The first component of the framework involves determining environmental conditions and selective pressures. The environment can be broadly understood as the aggregate of selective pressures encountered by a lineage (population or species), encompassing

features such as habitat, climate, biotic interactions, among others (Losos, 2011).

When focusing on repeated (convergent and parallel) phenotypic evolution, one should determine whether the derived phenotype results from similar environmental features. As an example, the mosaic distribution of an environment may result in repeated phenotypic evolution as organisms encounter similar environmental conditions and selective pressures while being geographically segregated (Barrett *et al.*, 2008; Konečná *et al.*, 2020; Sowersby *et al.*, 2021; Fulgione *et al.*, 2022). Additionally, transitions between environments can be biased (Baldwin *et al.*, 2021), and the biased and repeated occupation of habitats may lead to repeated phenotypic evolution. While we may intuitively think that repeated phenotypic evolution occurs only in the face of similar environmental conditions, this is not necessarily the case. For instance, the evolution of wings in insects, birds, pterodactyls and bats may have occurred due to different selective pressures: flying may enhance the ability of finding mates, or it may improve predator avoidance, or increase hunting efficiency. In a similar way, the repeated increase or decrease of body size in two lineages may result from predator avoidance or as a response to changes in temperature (Bergmann, 1848).

Natural selection driven by abiotic or biotic conditions may not necessarily be the only type of selection resulting in the evolution of in parallel or convergent phenotypes. For instance, shifts of body size or colouration in a given sex may result from sexual selection (Trillmich & Trillmich, 1984; Sanger *et al.*, 2013).

In sum, to fully understand parallel and convergent evolution, one should understand environmental conditions and the resulting selective pressures. Modern ecological tools such as species distributions models (Elith & Leathwick, 2009; White *et al.*, 2020) or hypervolumes (which represent ecological niches in an arbitrary number of environmental axes; (Blonder *et al.*, 2014; Blonder, 2018)), allow understanding of how environmental variables shape species' distributions and niches. For example, the overlap between hypervolumes may indicate that lineages share a given ecological niche, and this may be a strong indication of the role of environmental conditions in driving repeated phenotypic evolution. Another way to test the role of the environment is to quantify selective pressures directly, by using measurements of fitness and survival in an ecological setting (Losos, 2011; Cerca *et al.*, 2019).

Genotype

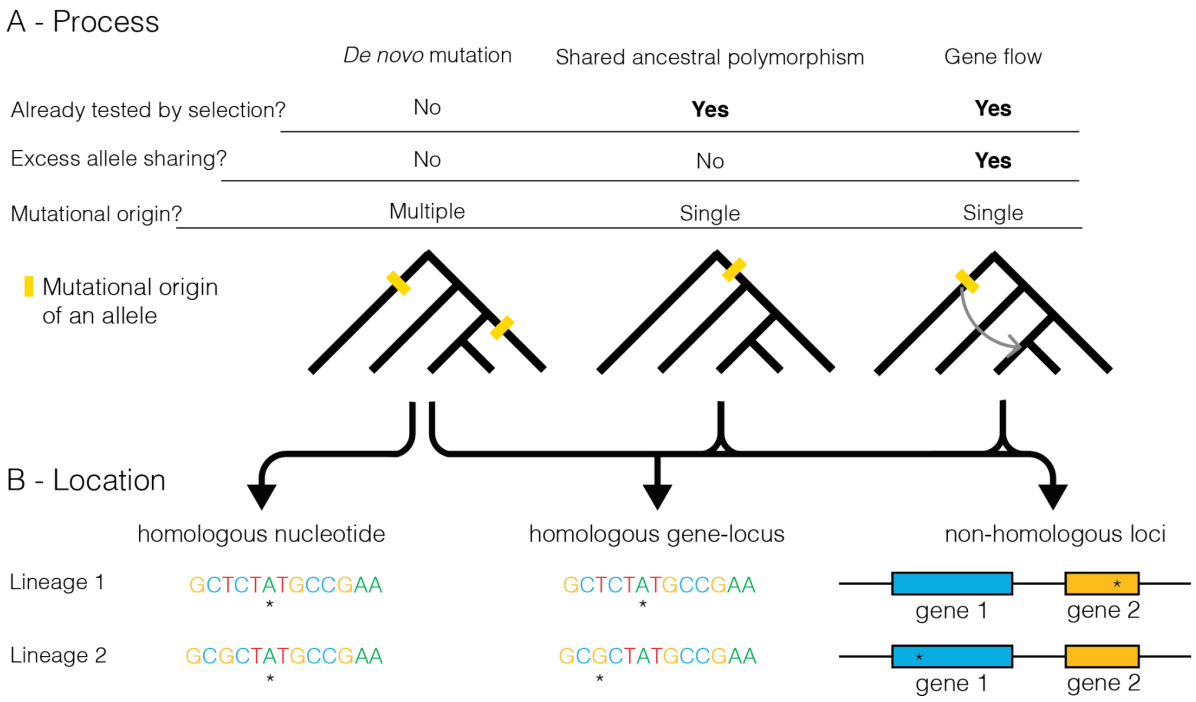

A - Process
De novo mutation
Shared ancestral polymorphism
Gene flow
Already tested by selection?
No
Yes
Yes
Excess allele sharing?
No
No
Yes
Mutational origin?
Multiple
Single
Single
Mutational origin of an allele
B - Location
homologous nucleotide
homologous gene-locus
non-homologous loci
Lineage 1
GCTCTATGCCGAA
*
GCTCTATGCCGAA
*
gene 1
gene 2
Lineage 2
GCGCTATGCCGAA
*
GCGCTATGCCGAA
*
gene 1
gene 2

Figure 2. Genomic basis of repeated (parallel or convergent) phenotypic evolution. A) *de novo* mutation takes place when derived phenotypes emerge from independent mutations, which have not yet been tested by selection, occur; Shared ancestral polymorphism or standing genetic variation occurs when the polymorphism has been tested by selection and is present in different lineages; gene flow can lead to phenotypic evolution when a given allele is recruited from another lineage through introgression. Gene flow can be detected by excess allele sharing. *De novo* mutations, shared ancestral polymorphisms and gene flow are not mutually exclusive (see Table 1). B) The processes can be further characterised based on its location: mutations at the same nucleotide (homologous nucleotide - only *de novo* mutation), at different nucleotides in the same gene (homologous locus - all three processes), and at different genes (non-homologous loci - all three processes). Mutations are marked with asterisks (*).

Table 1 Non-exhaustive evidence of overlap between genomic processes and their locations. Homologous nucleotide location is only considered under a scenario of *de novo* mutation.

System	Scope	Process			Location		
		De novo	Standing Genetic	Gene flow	At an homologous	At an homologous	At non-homologous

		mutation	Variation		nucleotide	gene-locus	nucleotides
Heliconius erato & H. melpomene	Adaptation to high-altitude (Montejo-Kovacevich *et al.*, 2021)	no	yes	yes	-	yes	yes
Drosophila ananassae (species group)	Male colouration (Signor *et al.*, 2016)	yes	yes	no	no	yes	yes
Wild tomato	Lineage diversification (Pease *et al.*, 2016)	yes	yes	yes	Not clear	yes	yes
Andean Waterfowl	Adaptation to high-altitude (Natarajan *et al.*, 2015)	yes	no	yes	no	yes	yes
Arabidopsis arenosa	Adaptation to serpentine soils (Konečná *et al.*, 2021)	yes	yes	no	no	yes	yes

Poeciliid fishes	Adaptation to hydrogen sulphide rich environments (Brown *et al.*, 2019)	yes	Plausible (or gene flow)	Plausible (or standing genetic variation)	no	yes	yes

The second component of the framework includes the organismal properties, which is split into the genotype and phenotype. The genomic basis of convergent and parallel phenotypic evolution is best understood when separating the ***(i)*** process (*De novo* mutation, shared ancestral polymorphism or standing genetic variation, and gene flow) (Stern, 2013), from the ***(ii)*** the location where it happens (homologous nucleotides, homologous gene-locus, and non-homologous gene-loci) (Figure 2).

De novo mutation refers to a change of the DNA sequence (Figure 2A). Notwithstanding that mutations are random, the 'random' nature of point mutations must be clarified before further discussion. Mutations are random in that it is not possible to predict their outcome as a single base pair can be deleted, duplicated, or converted to any other base pair. However, the occurrence and accumulation of mutations along the genome (i.e. the location of mutations) can be highly non-random, as there are areas in the genome with higher mutation rates (i.e. higher probability of a mutation to occur) and higher rates of mutation accumulation (i.e. higher accumulation of genetic polymorphisms). The increased accumulation of mutations at certain genomic locations is

affected by various processes such as pleiotropy, epistasis, epigenetics, the strength of selection, and population and size history (Stern & Orgogozo, 2009; Rosenblum *et al.*, 2014). This is well-represented by the high number of polymorphisms in animal mitochondria and sex-chromosomes partly due to their decreased effective population size. With this in mind, there are 'genomic hotspots of adaptive variation' where mutations are more likely to occur (Martin & Orgogozo, 2013). I suggest adding 'adaptive' as some genomic regions may have higher mutation rate, but these are not necessarily adaptive. For instance, in the case of microsatellites, higher mutation rates exist due to a higher probability of slippage by the polymerase. If a genomic hotspot of adaptive variation is present in the genome of two different lineages the likelihood of convergent or parallel phenotypic evolution may increase (Martin & Orgogozo, 2013). As an example, Xie and colleagues (Xie *et al.*, 2019) found that due to DNA fragility, there was a high occurrence of mutations in the enhancer of the Pitx1 gene. Mutations in the enhancer led to the parallel loss of pelvic hindfins in different three-spine stickleback populations.

The location of a particular mutation is therefore important when considering parallel and convergent phenotypic evolution. Following (Elmer & Meyer, 2011), I suggest the adoption of the terms ***(i)*** 'mutation at homologous nucleotide' when mutations occur at the same nucleotide; ***(ii)*** 'mutation at a homologous region' when mutations occur in different nucleotides at the same gene or extended genetic region (i.e. including promoters); ***(iii)*** 'mutation at non-homologous regions' when mutations occur in different genes (Figure 2B; (Elmer & Meyer, 2011). This

terminology is verbose, but is precise. It suits not only point mutations, but also larger mutational events, such as the insertion of transposable elements and inversions, as it is possible to map these events at the nucleotide level when aligning homologous regions of different genomes.

The second pattern to consider at the genomic level is shared ancestral polymorphism or standing genetic variation. The difference between *de novo* mutation and standing genetic variation is that the latter was already refined by selection, sometime in the past (Barrett & Schluter, 2008). Alternative alleles may exist in lineages because they are (or were) advantageous in a given environment, and, if these alleles are neutral or only slightly deleterious in the environment where the lineage currently exists, they can be maintained as low-frequency alleles (i.e. rare alleles) during many generations, especially in lineages with big population sizes (Barrett & Schluter, 2008; Sætre & Ravinet, 2019). Because these alleles have already been tested by selection, standing genetic variation can result in the repeated evolution of a phenotype, in a way faster (or more efficient) than *de novo* mutation (Barrett & Schluter, 2008; Gompel & Prud'homme, 2009; Bohutínská *et al.*, 2021).

The final genomic basis of repeated evolution to consider is gene flow. This evolutionary process is a powerful force of evolution as it can bring adaptive alleles across lineages. While the effect of gene flow can be multifarious (e.g. increase of genetic diversity, enhance or decrease the ability of a population to locally adapt), the introgression of alleles which were already tested by selection in another lineage may lead to

the parallel or convergent evolution of phenotypes (Pease *et al.*, 2016; Marques *et al.*, 2019b; Sowersby *et al.*, 2021).

In terms of the location, both gene flow and standing genetic variation occur at homologous and non-homologous gene-regions. Their effect is not observed at the nucleotide-level as these processes leave genomic-tracks, rather than single nucleotide changes (Martin *et al.*, 2014). I refer to 'gene-region' and not simply gene because phenotypic evolution is not necessarily linked to the coding region. For example, mutations on cis-regulatory regions led to the repeated evolution of phenotypes in species belonging to the *Drosophila* genus (Signor *et al.*, 2016).

These three genomic patterns may be non-exclusive, and the three were observed in a radiation of wild tomatoes (Table 1) (Pease *et al.*, 2016). Distinguishing between these processes has been historically challenging, but recent advancements in the field of genomics have greatly facilitated this task (Figure 2) (Lee & Coop, 2017). First, gene flow can be detected and quantified when analysing patterns of allele excess in a phylogenetic framework (Marques *et al.*, 2019b; Malinsky *et al.*, 2021). In specific, when considering a tree where P1 and P2 are sister species, and this clade is sister to P3 and an outgroup is added (as represented by the tree: (((P1, P2), P3), O);), the excess of allele-sharing between P3 and P1 or between P3 and P2 indicates gene flow between P3 and P1 or P2, respectively (Patterson's D or the ABBA-BABA statistic). The evidence of gene flow comes from the null hypothesis: under incomplete lineage sorting, similar ratios of ABBA (P3 and P2 share the B allele) and BABA (P3

and P1 share the B allele) are expected. A skew in the ABBA-BABA ratio points to the existence of gene flow. Other tests such as the F-branch statistic (where the tree is ((P1, P2),(P3, P4));) provide similar and complementary results. While ABBA-BABA and F-branch statistics have been some of the most popular ways to tease gene flow apart from other genomic scenarios, the study of genetic incompatibilities (i.e. genetic incompatibilities are more likely to occur in cases of gene flow than in cases of standing genetic variation), and long-range haplotypes (i.e. longer divergent haplotypes are expected to occur as a result of standing genetic variation) may also help distinguishing between scenarios (Marques *et al.*, 2019b).

The three main differences between standing genetic variation and *de novo* mutation are that: ***(i)*** the first has already been refined by selection (Barrett & Schluter, 2008), ***(ii)*** the mutation generally starts from a lower frequency (1/2N where N is the population size), and that ***(iii)*** to result in repeated phenotypic evolution, mutations must have multiple origins while standing genetic variation has a single origin. It is thus possible to distinguish between these two by employing a combination of analyses involving selective sweeps and coalescence (Ralph & Coop, 2015; Lee & Coop, 2017; Sætre & Ravinet, 2019). Consider a low-frequency allele (rare allele) that is picked up by natural selection. As natural selection increases the frequency of the allele, genomic regions which are physically close to that allele will also rise in frequency, hitchhiking with the selected variant - a phenomenon denominated as 'selective sweep'. The sweep will leave a 'V'-shaped signature on the physically linked polymorphisms: at the population-level, there will be a reduction

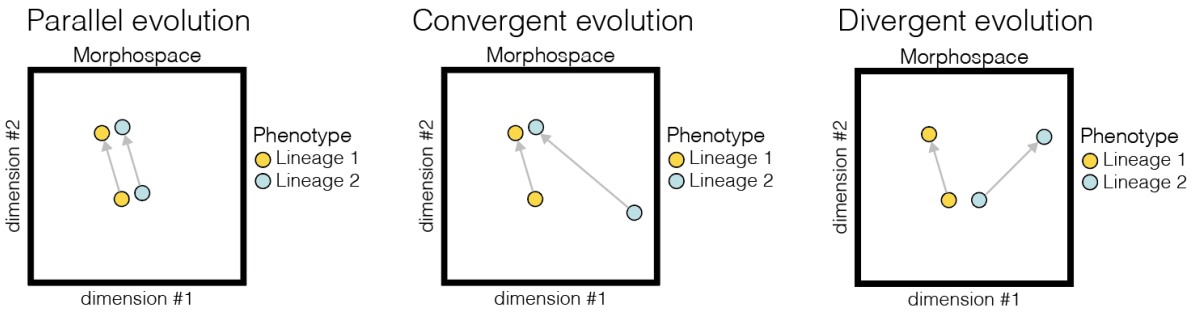
of polymorphisms in physiccally-close regions. Specifically, the closer polymorphisms are to the region picked up by selection, the more likely it is that they are removed from the population or have its allelic frequency decreased by the sweep. Distant polymorphisms will be less affected because of recombination, which breaks linkage between regions (Barrett & Schluter, 2008). Now, consider two scenarios where natural selection has the same strength. In one scenario, selection is acting on standing genetic variation, whereas in the second scenario on a *de novo* mutation. Because standing genetic variation has been present in the lineage for a given number of generations, there is an increased probability that recombination has generated different haplotypes. In other words, the allele that is being picked by selection is present with a more comprehensive representation of the catalog of polymorphisms of the population (Lee & Coop, 2017). The mutation scenario, on the other hand, entails that the mutation evolves in a single organism, thus having a starting frequency of 1/2N. Therefore, any closely-linked polymorphisms which are not present in that individual will likely be removed from the population (Hahn, 2018). To sum up, the reduction of closely-linked polymorphisms (strength of the sweep) is expected to be higher in the case of *de novo* mutation when contrasted with standing genetic variation, partly because of the starting frequency and the lack of recombination. Nonetheless, the evolutionary history of organisms can be complex and the interplay between selection and recombination may confound these effects. In such cases, employing simulations of the coalescent model or forward simulations, where users are able to specify scenarios (including the

evidence of gene flow, selection) and comparing these to the observed data may be of benefit (Lee & Coop, 2017; Nelson *et al.*, 2020). Finally, to lead to repeated phenotypic evolution, de novo mutation would have to occur at least twice, resulting in different sweeps.

A final, yet important consideration about these three patterns, concerns the 'repeated' nature of genome patterns, as all three can occur repeatedly, and result in repeated phenotypic evolution. *De novo* mutation can happen repeatedly (Konečná *et al.*, 2021; Fulgione *et al.*, 2022), especially in hotspots of adaptive variation, as discussed above (Xie *et al.*, 2019). Repeated events of gene flow have also been shown (Jones *et al.*, 2018; Giska *et al.*, 2019; Stolle *et al.*, 2022), and there can be a repeated recruitment of standing genetic variation in different lineages (Lai *et al.*, 2019; Konečná *et al.*, 2021). This, once again, highlights the need of coherent language.

Phenotype

A



B

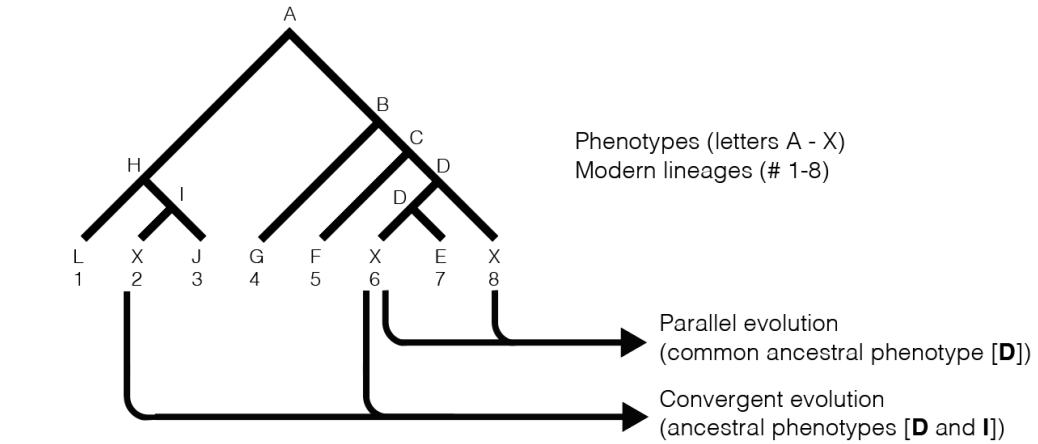


Figure 3 Parallel and convergent evolution at the phenotypic level. Parallel evolution involves the repeated evolution of a derived phenotype from a similar ancestral phenotype, whereas convergence involves the evolution of a derived phenotype from a dissimilar ancestral phenotype. A) Morphospace representation of parallel and convergent evolution. Divergent evolution is also represented for the purpose of comparison. The arrow provides directionality between ancestral and modern phenotypes. This follows the extensive framework by (Bolnick *et al.*, 2018). B) Phyletic representation of parallel and convergent phenotypic evolution. Phenotypes are represented by letters, and letters at internal branches represent the ancestral phenotypes (A-D, H, I).

Distinguishing between parallel or convergent phenotypic evolution is of fundamental interest to evolutionary biology. Parallel phenotypic evolution highlights that a derived phenotype can emerge from a similar ancestral starting point whereas convergent phenotypic evolution highlights that a derived phenotype can emerge from dissimilar starting points (Figure 3). As previously discussed, understanding of parallel and convergent phenotypic evolution benefits from the inclusion of genetic and ecological data, which allow drawing inferences on the drivers and the underlying basis of phenotypic evolution. To rigorously test hypotheses of parallel and convergent evolution, one needs to meticulously collect and analyse phenotypic data, and quantify the variation between derived phenotypes and to infer ancestral phenotypes. Modern tools such as principal component

analysis, and morphospaces allow decomposing phenotypic variation in axes of divergence (Figure 3A; e.g. (Cerca *et al.*, 2020)) thus allowing the estimation of phenotypic overlaps between different lineages (Bolnick *et al.*, 2018). The integration of ancestral characters, such as fossils or statistical reconstructions of ancestral states coupled with a phylogenetic backbone allows inferring patterns of phenotypic evolution, discerning between parallel, convergent and other forms of phenotypic evolution (Figure 3B; (Bolnick *et al.*, 2018; Pearce, 2020).

Finally, an important consideration pertains to changes in gene expression, constraints (developmental and genetic-architecture) and changes in amino-acid sequences. These should all be considered as phenotypic changes because, strictly speaking, these represent an expression of the genotype. However, recent evidence shows that parallel or convergent phenotypic traits can result from repeated and coherent alterations of gene expression (Vizueta *et al.*, 2019; Tobler *et al.*, 2021). I encourage the use of 'concordant changes in gene expression' when there is evidence for coherent changes in gene expression. I suggest this terminology because gene expression is dynamic and can change through time (Wos *et al.*, 2021).

The three-spined stickleback as an example of the utility of the framework

The best known case of parallel evolution is the three-spined stickleback, where marine populations have repeatedly expanded from marine environments to occupy freshwater environments such as

streams and lakes. The selective pressures resulting from freshwater environments have led to parallel phenotypic changes, such as the loss of the lateral plate armours, and alterations in pelvic spines, spine length, gill raker number and length, and colouration across the northern hemisphere (McGee *et al.*, 2013). However, the environmental heterogeneity of freshwater systems results in complex parallel and non-parallel phenotypic responses, highlighting the complex role of the environmental domain in shaping parallel phenotypic differentiation, and how only some traits ma evolve in parallel (Stuart *et al.*, 2017; Magalhaes *et al.*, 2021). While the first works found that parallel phenotypic evolution arose from standing genetic variation of low-frequency alleles, which are slightly deleterious in the marine environment, other populations have evolved parallel phenotypes by *de novo* mutation (Stern, 2013) or admixture (Marques *et al.*, 2019a). In some cases, parallel phenotypic changes have resulted from a combination of *de novo* mutation and standing genetic variation (Marques *et al.*, 2017), and nearly ~40% of the genome was repeatedly differentiated across lake-stream pair replicates (Rennison *et al.*, 2019). DNA fragility, mutability, genomic hotspots of adaptive variation and a certain degree of pleiotropy may contribute to the repeated differentiation of the genome (Chan *et al.*, 2010; Xie *et al.*, 2019; Rennison & Peichel, 2022), which ultimately underlie parallel phenotypic evolution.

The three-spined stickleback showcases the fine nuances of environmental and genomic features when studying parallel phenotypic evolution, which were highlighted by multidisciplinary efforts,

encompassing ecological, paleontological and genomic work. Stating that parallel selective pressures drive parallel genomic divergence, ultimately underlying the parallel evolution of traits leads to a loss of understanding due to the simplification and redundancy of the language.

Conclusions

The excitement for parallel and convergent phenotypic evolution draws ecologists, developmental biologists, geneticists, physiologists, systematists and even cancer biologists together (Supplementary Table 1). The popularity and interdisciplinary reach of these terms has likely contributed to the confusion observed in the literature. The framework herein proposed allows integrating the interdisciplinary nature of studying parallel and convergent evolution, and allows bypassing the difficulties identified (see introduction).

As the human species increases its impact on our planet, understanding the evolutionary potential of populations and species and the ability of species to repeatedly evolve in the face of certain environmental challenges is no longer of pure academic interest. In this context, parallel and convergent evolution studies are not only welcome, but increasingly pertinent. Getting lost in nuances will only slow us down.

Acknowledgements

The ideas, writing, and overall structure of this manuscript was greatly improved after suggestions and comments from Dimitar Dimitrov, Mafalda Sousa Ferreira, Siri Birkeland, Mark Ravinet, Andy

Foote, David Marques, Magdalena Bohutínská, Filip Kolář, and Cindy Santander. This manuscript is dedicated to my friends Filipe and Nicolas, who have repeatedly shown me that science became a productivity rat race, no longer meriting curiosity or originality - 'papers and grants'. This paper is dedicated to them, and every single brilliant scientist leaving academia for similar reasons. The ideas of this paper came during a short stay in Ayampe, Ecuador, while I was funded by the Norwegian Research Council (project 287327). I thank Tora for her unconditional and constant encouragement. Without your support this would not have been possible.

Author Contributions

JC did everything.

References

Arendt, J. & Reznick, D. 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol. Evol.* **23**: 26–32.

Baldwin, B.G., Wood, K.R. & Freyman, W.A. 2021. Directionally biased habitat shifts and biogeographically informative cytonuclear discordance in the Hawaiian silversword alliance (Compositae). *Am. J. Bot.* **108**: 2015–2037.

Barrett, R.D.H., Rogers, S.M. & Schluter, D. 2008. Natural selection on a major armor gene in threespine stickleback. *Science* **322**: 255–257.

Barrett, R.D.H. & Schluter, D. 2008. Adaptation from standing genetic

variation. *Trends Ecol. Evol.* **23**: 38–44.

Bergmann, C. 1848. *Über die Verhältnisse der Wärmeökonomie der Thiere zu ihrer Größe*.

Blonder, B. 2018. Hypervolume concepts in niche- and trait-based ecology. *Ecography* **41**: 1441–1455. Wiley.

Blonder, B., Lamanna, C., Violle, C. & Enquist, B.J. 2014. Then-dimensional hypervolume. *Glob. Ecol. Biogeogr.* **23**: 595–609. Wiley.

Bohutínská, M., Vlček, J., Yair, S., Laenen, B., Konečná, V., Fracassetti, M., *et al.* 2021. Genomic basis of parallel adaptation varies with divergence in Arabidopsis and its relatives. *Proc. Natl. Acad. Sci. U. S. A.* **118**.

Bolnick, D.I., Barrett, R.D.H., Oke, K.B., Rennison, D.J. & Stuart, Y.E. 2018. (non)parallel evolution. *Annu. Rev. Ecol. Evol. Syst.* **49**: 303–330. Annual Reviews.

Brown, A.P., McGowan, K.L., Schwarzkopf, E.J., Greenway, R., Rodriguez, L.A., Tobler, M., *et al.* 2019. Local ancestry analysis reveals genomic convergence in extremophile fishes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **374**: 20180240.

Cerca, J., Agudo, A.B., Castro, S., Afonso, A., Alvarez, I. & Torices, R. 2019. Fitness benefits and costs of floral advertising traits: insights from rayed and rayless phenotypes of Anacyclus (Asteraceae). *Am. J. Bot.* **106**: 231–243.

Cerca, J., Meyer, C., Stateczny, D., Siemon, D., Wegbrod, J., Purschke, G., *et al.* 2020. Deceleration of morphological evolution in a cryptic species complex and its link to paleontological stasis. *Evolution* **74**: 116–131.

Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal, G., Jr, Shapiro, M.D., Brady, S.D., *et al.* 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* **327**: 302–305.

Colosimo, P.F., Hosemann, K.E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., *et al.* 2005. Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles.

Elith, J. & Leathwick, J.R. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time.

Elmer, K.R. & Meyer, A. 2011. Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends Ecol. Evol.* **26**: 298–306.

Fulgione, A., Neto, C., Elfarargi, A.F., Tergemina, E., Ansari, S., Göktay, M., *et al.* 2022. Parallel reduction in flowering time from de novo mutations enable evolutionary rescue in colonizing lineages. *Nat. Commun.* **13**: 1461.

Giska, I., Farelo, L., Pimenta, J., Seixas, F.A., Ferreira, M.S., Marques, J.P., *et al.* 2019. Introgression drives repeated evolution of winter coat color polymorphism in hares. *Proc. Natl. Acad. Sci. U. S. A.* **116**: 24150–24156.

Gompel, N. & Prud'homme, B. 2009. The causes of repeated genetic

evolution. *Dev. Biol.* Elsevier.

Hahn, M.W. 2018. *Molecular Population Genetics*. Oxford University Press.

Harrison, R.G. 2012. The language of speciation. *Evolution* **66**: 3643–3657.

Jones, M.R., Mills, L.S., Alves, P.C., Callahan, C.M., Alves, J.M., Lafferty, D.J.R., *et al.* 2018. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science* **360**: 1355–1358.

Konečná, V., Bray, S., Vlček, J., Bohutínská, M., Požárová, D., Choudhury, R.R., *et al.* 2021. Parallel adaptation in autopolyploid Arabidopsis arenosa is dominated by repeated recruitment of shared alleles. *Nat. Commun.* **12**: 4979.

Konečná, V., Yant, L. & Kolář, F. 2020. The Evolutionary Genomics of Serpentine Adaptation. *Front. Plant Sci.* **11**: 574616.

Lai, Y.-T., Yeung, C.K.L., Omland, K.E., Pang, E.-L., Hao, Y., Liao, B.-Y., *et al.* 2019. Standing genetic variation as the predominant source for adaptation of a songbird. *Proc. Natl. Acad. Sci. U. S. A.* **116**: 2152–2157.

Leander, B. 2008. Different modes of convergent evolution reflect phylogenetic distances: a reply to Arendt and Reznick.

Lee, J.-H., Lewis, K.M., Moural, T.W., Kirilenko, B., Borgonovo, B., Prange, G., *et al.* 2018. Molecular parallelism in fast-twitch muscle proteins in echolocating mammals. *Sci Adv* **4**: eaat9660.

Lee, K.M. & Coop, G. 2017. Distinguishing Among Modes of Convergent

Adaptation Using Population Genomic Data. *Genetics* **207**: 1591–1619.

Losos, J.B. 2010. Adaptive radiation, ecological opportunity, and evolutionary determinism. American Society of Naturalists E. O. Wilson award address. *Am. Nat.* **175**: 623–639.

Losos, J.B. 2011. Convergence, adaptation, and constraint. *Evolution* **65**: 1827–1840.

Magalhaes, I.S., Whiting, J.R., D'Agostino, D., Hohenlohe, P.A., Mahmud, M., Bell, M.A., *et al.* 2021. Intercontinental genomic parallelism in multiple three-spined stickleback adaptive radiations. *Nat Ecol Evol* **5**: 251–261.

Malinsky, M., Matschiner, M. & Svardal, H. 2021. Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Mol. Ecol. Resour.* **21**: 584–595. Wiley.

Marques, D.A., Lucek, K., Sousa, V.C., Excoffier, L. & Seehausen, O. 2019a. Admixture between old lineages facilitated contemporary ecological speciation in Lake Constance stickleback. *Nat. Commun.* **10**: 4240.

Marques, D.A., Meier, J.I. & Seehausen, O. 2019b. A Combinatorial View on Speciation and Adaptive Radiation. *Trends Ecol. Evol.* **34**: 531–544.

Marques, D.A., Taylor, J.S., Jones, F.C., Di Palma, F., Kingsley, D.M. & Reimchen, T.E. 2017. Convergent evolution of SWS2 opsin facilitates adaptive radiation of threespine stickleback into different light environments. *PLoS Biol.* **15**: e2001627.

Martin, A. & Orgogozo, V. 2013. The Loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution* **67**: 1235–1250.

Martin, S.H., Davey, J.W. & Jiggins, C.D. 2014. Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. *Mol. Biol. Evol.* **32**: 244–257. Oxford Academic.

McGee, M.D., Schluter, D. & Wainwright, P.C. 2013. Functional basis of ecological divergence in sympatric stickleback. *BMC Evol. Biol.* **13**: 277.

Mendes, F.K., Hahn, Y. & Hahn, M.W. 2016. Gene Tree Discordance Can Generate Patterns of Diminishing Convergence over Time. *Mol. Biol. Evol.* **33**: 3299–3307.

Montejo-Kovacevich, G., Meier, J.I., Bacquet, C.N., Warren, I.A., Chan, Y.F., Kucka, M., *et al.* 2021. Repeated genetic adaptation to high altitude in two tropical butterflies.

Natarajan, C., Projecto-Garcia, J., Moriyama, H., Weber, R.E., Muñoz-Fuentes, V., Green, A.J., *et al.* 2015. Convergent Evolution of Hemoglobin Function in High-Altitude Andean Waterfowl Involves Limited Parallelism at the Molecular Sequence Level. *PLoS Genet.* **11**: e1005681.

Nelson, D., Kelleher, J., Ragsdale, A.P., Moreau, C., McVean, G. & Gravel, S. 2020. Accounting for long-range correlations in genome-wide simulations of large cohorts. *PLoS Genet.* **16**: e1008619.

Pearce, T. 2020. Convergence and parallelism in evolution: a neo-Gouldian account. *Br. J. Philos. Sci.* The University of Chicago Press.

Pease, J.B., Haak, D.C., Hahn, M.W. & Moyle, L.C. 2016. Phylogenomics Reveals Three Sources of Adaptive Variation during a Rapid Radiation. *PLoS Biol.* **14**: e1002379.

Ralph, P.L. & Coop, G. 2015. Convergent Evolution During Local Adaptation to Patchy Landscapes. *PLoS Genet.* **11**: e1005630.

Rennison, D.J. & Peichel, C.L. 2022. Pleiotropy facilitates parallel adaptation in sticklebacks. *Mol. Ecol.* **31**: 1476–1486.

Rennison, D.J., Stuart, Y.E., Bolnick, D.I. & Peichel, C.L. 2019. Ecological factors and morphological traits are associated with repeated genomic differentiation between lake and stream stickleback. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **374**: 20180241.

Rosenblum, E.B., Parent, C.E. & Brandt, E.E. 2014. The Molecular Basis of Phenotypic Convergence. , doi: 10.1146/annurev-ecolsys-120213-091851. Annual Reviews.

Sætre, G.-P. & Ravinet, M. 2019. *Evolutionary Genetics: Concepts, Analysis, and Practice*. Oxford University Press.

Sanger, T.J., Sherratt, E., McGlothlin, J.W., Brodie, E.D., 3rd, Losos, J.B. & Abzhanov, A. 2013. Convergent evolution of sexual dimorphism in skull shape using distinct developmental strategies. *Evolution* **67**:

2180–2193.

Signor, S.A., Liu, Y., Rebeiz, M. & Kopp, A. 2016. Genetic Convergence in the Evolution of Male-Specific Color Patterns in Drosophila. *Curr. Biol.* **26**: 2423–2433.

Sowersby, W., Cerca, J., Wong, B.B.M., Lehtonen, T.K., Chapple, D.G., Leal-Cardín, M., *et al.* 2021. Pervasive admixture and the spread of a large-lipped form in a cichlid fish radiation. *Mol. Ecol.* **30**: 5551–5571.

Stayton, C.T. 2015. The definition, recognition, and interpretation of convergent evolution, and two new measures for quantifying and assessing the significance of convergence. *Evolution* **69**: 2140–2153.

Stern, D.L. 2013. The genetic causes of convergent evolution. *Nat. Rev. Genet.* **14**: 751–764.

Stern, D.L. & Orgogozo, V. 2009. Is genetic evolution predictable? *Science* **323**: 746–751.

Stolle, E., Pracana, R., López-Osorio, F., Priebe, M.K., Hernández, G.L., Castillo-Carrillo, C., *et al.* 2022. Recurring adaptive introgression of a supergene variant that determines social organization. *Nat. Commun.* **13**: 1180.

Stoltzfus, A. & McCandlish, D.M. 2017. Mutational Biases Influence Parallel Adaptation. *Mol. Biol. Evol.* **34**: 2163–2172.

Stuart, Y.E., Veen, T., Weber, J.N., Hanson, D., Ravinet, M., Lohman, B.K., *et al.* 2017. Contrasting effects of environment and genetics generate a

continuum of parallel evolution.

Szukala, A., Lovegrove-Walsh, J., Luqman, H., Fior, S., Wolfe, T.M., Frajman, B., *et al.* 2022. Polygenic routes lead to parallel altitudinal adaptation in Heliosperma pusillum (Caryophyllaceae). *Mol. Ecol.*, doi: 10.1111/mec.16393.

Tobler, M., Greenway, R. & Kelley, J.L. 2021. Ecology drives the degree of convergence in the gene expression of extremophile fishes.

Trillmich, F. & Trillmich, K.G.K. 1984. The mating systems of pinnipeds and marine iguanas: convergent evolution of polygyny. *Biol. J. Linn. Soc. Lond.* **21**: 209–216. Oxford University Press (OUP).

Vizueta, J., Macías-Hernández, N., Arnedo, M.A., Rozas, J. & Sánchez-Gracia, A. 2019. Chance and predictability in evolution: The genomic basis of convergent dietary specializations in an adaptive radiation. *Mol. Ecol.* **28**: 4028–4045.

Wellborn, G.A. & Langerhans, R.B. 2015. Ecological opportunity and the adaptive diversification of lineages. *Ecol. Evol.* **5**: 176–195.

White, O.W., Reyes-Betancort, J.A., Chapman, M.A. & Carine, M.A. 2020. Geographical isolation, habitat shifts and hybridisation in the diversification of the Macaronesian endemic genus Argyranthemum (Asteraceae). *New Phytol.* **228**: 1953–1971.

Wood, T.E., Burke, J.M. & Rieseberg, L.H. 2005. Parallel genotypic adaptation: when evolution repeats itself. *Genetica* **123**: 157–170.

Wos, G., Bohutínská, M., Nosková, J., Mandáková, T. & Kolář, F. 2021. Parallelism in gene expression between foothill and alpine ecotypes in Arabidopsis arenosa. *Plant J.* **105**: 1211–1224.

Xie, K.T., Wang, G., Thompson, A.C., Wucherpfennig, J.I., Reimchen, T.E., MacColl, A.D.C., *et al.* 2019. DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. *Science* **363**: 81–84.